

-

邮箱：[yongjieliu@ir.hit.edu.cn](mailto:yongjieliu@ir.hit.edu.cn)

电话：15546439752

主页：<http://ir.hit.edu.cn/~yongjieliu/>

* 语言：英语六级、日语二级
* 曾获奖项： 阿里天池健康AI复赛：70/3152 国家奖学金

一等奖学金 企业奖学金(安然纳米) 2017届优秀毕业生

**奖项**

**项目一：大词林（www.bigcilin.com） 2016.09-至今**

* 主要工作：使用动态加密技术更新了“大词林”的防爬模块；融入实体同义词信息，完善了“大词林”上位词查询功能等。目前正从百度百科爬取信息，补充“大词林”横向关系模块；对“大词林”的架构及其构建过程较为熟悉
* 1.收集词语先验概率、是否标题等多种特征，使用SVM对该词语是否为实体上位词进行二分类

2.提出合理猜想以自动获取SVM训练语料

**项目二：阿里天池——美年健康AI大赛 复赛60/3152 2018.05-2018.06**

* 主要技术：lightGBM、GBDT、特征工程
* 1.数据预处理，出去稀疏数据，将简单文本特征编码成分类型特征，对数值型和分类型数据进行了正态化等多项预处理，使用 LSTM+CRF 对复杂文本进行序列标注，并从中抽取出一系列关系作为特征，然后根据通过 GBDT 得到的特征重要性对特征进行降维。模型方面使用 lightgbm+xgboost 等模型作为底层模型，Gradient Boosting 回归作为上层模型的 stacking 架构。

**论文一：Effective Deep Memory Networks for Distant Supervised Relation Extraction IJCAI 2017 第五作者 负责实验部分 2016.11-2017.02**

* 使用框架：theano 主要技术：深度学习、word attention、relation attention
* 1.CNN抽取单句特征形成embedding表示时，加入句子中不同词对当前实体对的影响

2.将笼统的多分类替换为针对某一类别的二分类

3.多个句子embedding合成entity pair表示时，加入每个句子对不同类别的分类影响

**论文二：融入搜索引擎的启发式开放域三元组抽取**

**CCKS 2017 第一作者 2017.03-2017.07**

* 1.利用LTP(https://www.ltp-cloud.com/demo/)句法分析工具，从满足定中、主谓宾、主谓动补、定语后置的句法中抽取候选三元组

2.利用搜索引擎搜索实体间可能的关系词，利用tf\*idf筛选关系词候选，利用词向量计算关系词与关系候选词的相似度作为置信度，进一步过滤

**项目经历**

**2013.09—2017.06 本科 哈尔滨工业大学 软件工程**

**2017.09— 至今 研究生 哈尔滨工业大学 计算机技术**

.

**教育背景**

年龄：23(2019年毕业)

学历：硕士

–

**刘勇杰**